

## INCONGRUITY IN 360-DEGREE FEEDBACK RATINGS AND COMPETING MANAGERIAL VALUES: EVIDENCE FROM A PUBLIC AGENCY SETTING

SHAHIDUL HASSAN AND JOHN ROHRBAUGH  
*UNIVERSITY AT ALBANY, SUNY*

**ABSTRACT:** *This study examined the sources, patterns, and implications of incongruence in 360-degree feedback ratings in public organizations using the Competing Values Framework for organizational effectiveness. Performance self-ratings from 68 high-performing, middle-level public sector managers, as well as parallel ratings provided by their supervisors, peers, and subordinates, were assembled and analyzed. Results indicated that rating incongruence existed across organizational roles and resulted from raters' unique role perspectives. Implications of incongruence in 360-degree feedback ratings with respect to developing effective systems for management development in public organizations are discussed in detail.*

---

### INTRODUCTION

The performance effectiveness of managerial work is highly complex and ambiguous in public organizations (Rainey 1997; 1989; Pandey and Wright 2006). Many factors contribute to this complexity and ambiguity, including the variety of roles, often conflicting, public sector managers need to perform in their jobs, the number of constituents and stakeholders to whom they need to attend every day, and the organizational, social, and political structures embedded in their work environment. In such a context, performance feedback plays an important role in guiding managers' behavior on the job. Performance feedback provides managers with an accurate sense of how others perceive their performance and helps them to identify both their strengths and their weaknesses. In addition, feedback is an essential component in managers' self-regulation process (Ashford and Tsui 1991), allowing them to detect discrepancies in their behaviors and subsequently to select appropriate performance-enhancing strategies such as adjusting their goals and/or their behaviors (Locke 1968; Locke and Latham 1990).

Managers receive performance feedback in many ways and from many sources. Of course, feedback may be inherent in a managerial task itself (Ashford and Tsui 1991), and managers also may be able to infer the effectiveness of their performance from indirect and informal cues (Ashford and Tsui 1991). However, direct and useful feedback is not always readily available in public organizations without the implementation of a formal system. Public organizations traditionally have used a top-down approach to assess managers' performance, but recently both researchers and practitioners have criticized such systems for inadequately assisting public sector managers in clearly identifying their strengths and weaknesses so as to improve their performance (deLeon and Ewen 1997; Kellough and Selden 1997; Pynes 1997; Selden, Ingraham, and Jacobson 2001). Current innovations in performance management systems in both private and public sector organizations encourage employees to collaborate in identifying and prioritizing their organizational performance goals and then aligning them with performance expectations (Selden et al. 2001). The aim is to develop a culture of improved performance in organizations using employee participation, collaboration, and empowerment. At a minimum, any effective performance management system in public organizations should provide managers with multiple sources of performance feedback, as well as opportunities to develop managerial skills and competencies through training and mentoring in areas where they exhibit weakness (Selden et al. 2001).

Multi-source feedback, commonly known as 360-degree feedback, is an innovative management development technique that can assist public organizations to develop an effective employee performance management system (deLeon and Ewen 1997). Whereas traditional performance evaluation methods used in public organizations provide only single-source assessments, typically from immediate supervisors, 360-degree feedback involves evaluations from supervisors, peers, subordinates, and sometimes even customers, in addition to managers' own self-assessments (Bracken et al. 2001). In addition, ratings in 360-degree feedback usually are made anonymously—with the exception of the immediate supervisor—and not accompanied by face-to-face discussion (Mount et al. 1998; Bracken et al. 2001). These procedures are followed to increase the likelihood that raters provide feedback that is more candid and, hence, more beneficial to managers, as well as to ensure that managers can effectively use feedback from various sources in a non-threatening setting.

The reasoning in support of a multi-source approach is that those who have frequent interactions with managers, who are fairly knowledgeable about their performance, and whose opinions are highly valued by them are in the best position to provide the kind of information that they need to improve their performance (Mount et al. 1998). This line of thinking also is characterized by a shift away from traditional hierarchical structures observed in public organizations to matrix structures and team-based organization of work, where feedback from multiples sources is thought to be central to improve managers' effectiveness. Given this perspective, 360-degree feedback can be an empowering mechanism in public organizations, because it gives both peers and subordinates a voice in assessing how effective their managers or colleagues are as leaders.

The growing use of 360-degree feedback in organizations over the past two decades is attributable to a number of factors, including the dramatic rise of formal leadership and management development programs; changes in work arrangements from traditional hierarchies to flatter, cross-functional structures; and the appeal of using quantifiable performance measures for assessing managerial effectiveness (Toegel and Conger 2003). However, the popularity of 360-degree feedback has led to organizational applications well apart from management development, including its use for routine performance appraisals that can have potentially negative consequences for individual managers (Toegel and Conger 2003). Designed primarily as a system for management development, 360-degree feedback had not been intended for use as a decision support tool for promotions, dismissals, or compensation (Mount et al. 1998; Bracken et al. 2001; Waldman, Atwater, and Antonioni 1998). In fact, there is a growing concern among researchers about the organizational damage that may be incurred by extending the use of 360-degree feedback as part of a performance appraisal system (Dalton 1997; deLeon and Ewen 1997; DeNisi and Klunger 2000; Pollack and Pollack 1996; Toegel and Conger 2003). This concern arises from the apprehension that, if 360-degree feedback is used for both performance appraisal and management development, then the management development process may be perceived as potentially punitive, one that forces rather than enables change (Pollman 1997). Further, a key goal of 360-degree feedback has been to foster continuous learning by managers (Toegel and Conger 2003). To achieve this goal, a psychologically safe environment is essential in which managers can comfortably provide and receive genuine feedback (Dalton 1997).

Research on 360-degree feedback mainly has focused on the psychometric properties of multi-source ratings, especially on the nature, sources, and practical implications of rating incongruence. Rating incongruence is defined as the degree to which ratings from multiple sources are dissimilar to each other (Mersman and Donaldson 2000). Although rating incongruence in 360-degree feedback exists for valid reasons and could be well integrated in the process of management development (Borman 1997; Hooijberg and Choi 2000; Salam, Cox, and Sims 1997; Tornow 1993), a high degree of incongruence between others' ratings of a manager (usually referred as "other-other" rating incongruence) generally has been considered undesirable. This is because, to guide behavioral change effectively, managers need to receive focused and coherent information about the areas where they should improve their performance (Borman 1997). Managers with inconsistent patterns of feedback from their supervisors, peers, and subordinates may find the feedback confusing and unhelpful and, therefore, may become less motivated to rely on it for improving their performance (Miller and Cardy 2002).

Since traditional methods of performance appraisal have been structured in accord with an implicit desire to minimize incongruence across ratings to increase the apparent validity of the evaluation process (Salam et al. 1997), incongruence between multiple sources of feedback often has been considered undesirable and a fundamental component of rating error. Nevertheless, theoretically integrated empirical explorations of the sources of rating incongruence remain quite limited. This gap provided the impetus for the current study, including the following key

research questions: (1) What factors may influence raters' assessment of managers' performance that contribute to rating incongruence? (2) Does rating incongruence exist uniformly in different managerial performance domains? and (3) How can a better understanding of rating incongruence improve the use of 360-degree feedback for human resource development in public organizations?

## LITERATURE REVIEW AND RESEARCH HYPOTHESES

Although other–other rating incongruence in 360-degree feedback has been investigated, the extent of incongruity has not been found to be large; such results have encouraged its growing use. Conway and Huffcutt (1997) undertook a meta-analysis of multi-source feedback studies completed since 1950 in which managers had been rated by both their subordinates and supervisors, peers and supervisors, and/or subordinates and peers. Mean between-source correlations (corrected using within-source interrater reliabilities) were as high as .80 for peer-supervisor rating pairs and no lower than .57 for subordinate-supervisor rating pairs. Thus, there appears to be considerable consensus that supervisor, peer, and subordinate ratings (frequently identified as “other–other” ratings) can be expected to be somewhat in parallel with each other.

Less consistent, however, is the agreement between managers' self-ratings and the ratings of others (frequently identified as “self–other” ratings). Conway and Huffcutt (1997) reported mean between-source correlations narrowly between .26 and .29 for self–peer, self–subordinate, and self–supervisor rating pairs. An earlier meta-analysis conducted by Harris and Schaubroeck (1988) produced similar results with corrected correlations between .27 and .31; comparable results have been found in more recent studies (Atkins and Wood 2002; Atwater et al. 2005; Hooft, Henk, and Minne 2006). Given this compilation of evidence, the first two hypotheses for the present study appeared to be obvious assertions to test with a reasonable expectation that they would find support:

H1: There will be incongruence between self–other ratings.

H2: The degree of incongruence between self–other ratings will be higher than that between other–other ratings.

The lack of convergence between self and other ratings has led researchers to probe the sources of rating incongruence. Much research has shown that self-perceptions play an important role in individuals' own performance evaluation and that self-perceptions are influenced by a variety of social, personality, and cognitive factors (Ashford 1989; Yammarino and Atwater 1993; Warr and Bourne 1999). For example, research has shown that self-ratings can be affected by self-esteem (Yammarino and Atwater 1993), leniency (Fahr and Dobbins 1989; Fox and Dinur 1988), social desirability (Podsakoff and Organ 1986), and attributional biases (DeVader, Bateson, and Lord 1986). Since self-assessment can be a threatening situation for many individuals, cognitive factors such as self-protection (Wohlers and London 1989), self-concept maintenance (Atwater and Yammarino 1992), and self-presentation pressures (Ashford 1989) can induce managers to

overestimate their own performance in an effort to maintain a positive self-image. Such pressures, of course, can lead to self–other rating incongruence, especially when 360-degree feedback ratings are used in systems of performance appraisal rather than in management development programs (Bracken et al. 2001; Waldman, Atwater, and Antonioni 1998).

A variety of studies have shown that situations also arise in which managers underestimate—rather than overestimate—their own performance. Self-ratings that underestimate managers' own performance lead to incongruity in self–other ratings, as well. Managers who underestimate their own performance have been found to be more effective than those who appear to inflate their self-ratings. Van Velsor, Taylor, and Leslie (1993), for example, found that overraters (i.e., those with self-ratings above other-ratings) received the lowest subordinate ratings on managerial roles compared to underraters (i.e., those with self-ratings below other-ratings). Similarly, Atwater et al. (1998) examined multi-source feedback in a large sample of private sector managers and found that effectiveness was the lowest for managers who overestimated their performance. They also discovered that the effectiveness of managers was highest when self and other ratings were all fairly high, and when managers' self-ratings were significantly lower than the ratings of others. These results are consistent with current socio-psychological research indicating that highly competent people will underestimate their own performance (Burson, Larrick, and Klayman 2006; Kruger and Dunning 1999, 2000).

Many scholars (Borman 1997; Bozeman 1997; Hooijberg and Choi 2000; Murphy and Cleveland 1995; Salam et al. 1997; Tornow 1993) have suggested that rating incongruence also can occur because raters from varying organizational levels evaluate alternative aspects of a manager's performance and/or have access to different information. Multiple perspectives, even when apparently contradictory, are not necessarily in error but provide complementary insights (Salam et al. 1997; Tornow 1993). These researchers have suggested that the utility of multi-source feedback actually derives from understanding the nature of rating differences observed across rater levels and interpreting them accurately to guide managers' behavioral change (Borman 1997; Mersman and Donaldson 2000). Therefore, one should not automatically expect nor desire a high degree of rating congruence (Borman 1997; Mersman and Donaldson 2000). In fact, a high degree of rating congruence may be undesirable as it would indicate that additional ratings only offer redundant information and collecting them would be a waste of organizational resources (Borman 1997; Mersman and Donaldson 2000).

Borman (1997) has summarized well the relevance of understanding the unique perspectives of raters in order to clarify the sources of rating incongruence. Raters may observe different managerial behaviors, employ different weighting of criteria, and use different interpretive systems. The factor that appears to be particularly critical in this perspective is the managerial roles each rater would consider important and relevant. Managers assume a wide range of responsibilities associated with their positions. Those who observe and evaluate them, including the managers' own self-reflections, may adopt different perspectives about what attributes are most relevant and important. Vandenberg, Lance, and Taylor (1997, 29) stated that “our

personal conceptualization of what constitutes performance is largely a function our social perceptions of the individual being rated, and our interaction goals with that individual due to our position in the organization.” From this perspective, the relative organizational position of the raters will have a significant bearing on the managerial attributes they select and emphasize, as well as affect the strictness or leniency of their judgments.

Several studies provide empirical support for such reasoning. While examining 360-degree feedback data for 456 professionals in a public agency, Pulakos, Schmitt, and Chan (1996) discovered a significant difference in the fit of the performance effectiveness models of peers and supervisors. In a longitudinal field study, Salam and others (1997) found that leaders who were seen as challenging the status quo and encouraging subordinates’ independent action were rated lower by their supervisors but rated higher by their subordinates. They concluded that the lack of congruence in ratings from different sources was not necessarily an error but stemmed from legitimate differences in the organizational role of the person rating the manager. Hooijberg and Choi (2000) examined the extent to which raters differed in the leadership roles that they associated with leadership effectiveness. They found that, depending on managers’ relationships with their raters (i.e., supervisor, peer, or subordinate), different leadership roles were connected with leadership effectiveness. Hooijberg and Choi (2000, 342) concluded that, “if organizational constituents differ fundamentally in the leadership roles they associate with effectiveness, 360-degree feedback researchers should not only pay attention to rater agreement, but also to substantive differences.”

Although various typologies have been proposed for clarifying the responsibilities embedded in managerial work (Fayol 1916; Gulick 1937; Mintzberg 1973), a more contemporary model for considering the leadership roles associated with organizational effectiveness is the Competing Values Framework (Quinn and Rohrbaugh 1983). The Competing Values Framework (CVF) originally was proposed and applied to study organizational effectiveness (Rohrbaugh 1983), but later extensions of the model moved beyond the organizational level of analysis to studies at both the group (McCartt and Rohrbaugh 1995; Reagan and Rohrbaugh 1990) and individual levels (Denison, Hooijberg, and Quinn 1995; Hooijberg and Choi 2000). Far more integrative than merely a list of various roles, the CVF suggests that the key criteria for managerial effectiveness can be organized in a three-dimensional space specified by competing organizational values. The first value dimension is related to organization focus, from an internal emphasis on the well-being and development of people in the organization to an external emphasis on the well-being and development of the organization itself. The second value dimension is related to structure, from an emphasis on achieving flexibility to an emphasis on achieving control. The third value dimension is related to means and ends, from an emphasis on critical mechanisms and processes to an emphasis on consequences and outcomes. The relationships between these three sets of value dimensions and the eight effectiveness criteria are shown in Figure 1.

Rohrbaugh (2005) has pointed out that the CVF reaffirms an important facet of Parson’s now-classic theory (1959) of collectivities or systems of action. In particular,

<b>INTERNAL FOCUS</b>	<b>FLEXIBILITY</b>		<b>EXTERNAL FOCUS</b>
	<b><u>Human Relations Model</u></b>	<b><u>Open Systems Model</u></b>	
	Means: Cohesion, morale Ends: Human resource development	Means: Flexibility, readiness Ends: Growth, resource acquisition	
	Means: Information management, communication Ends: Stability, control <b><u>Internal Processes Model</u></b>	Means: Planning, goal setting Ends: Productivity, efficiency <b><u>Rational Goals Model</u></b>	
<b>CONTROL</b>			

**Figure 1.** The Competing Values Framework.

the CVF quadrants can be seen to match Parson's specification of four functional prerequisites for sustaining systems: (1) "pattern maintenance": the human relations model; (2) "adaptation": the open systems model; (3) "goal attainment": the rational goal model; and (4) "integration": the internal process model. The human relations model emphasizes flexibility, has an internal focus, and stresses the achievement of human resource development through increased cohesion and morale. The open systems model emphasizes flexibility, has an external focus, and stresses the achievement of resource acquisition and growth through greater adaptability and readiness. The rational goals model emphasizes control, has an external focus, and stresses the achievement of productivity and efficiency through better planning and goal setting. The internal process model emphasizes control, has an internal focus, and stresses the achievement of operational stability through enhanced communication and information management.

The quadrants of the CVF offer a potentially useful model for considering the issues surrounding rater incongruence in 360-degree feedback. Because managers' supervisors, subordinates, and peers occupy a variety of positions in an organizational hierarchy with the attendant differences in their roles and responsibilities, as raters they might be expected to sample different behavior, use different cues, and apply different weighting systems in making performance evaluations (Borman 1997). Without taking the four performance domains of the CVF into account in any study of multi-source feedback, the basis of rater incongruence may not be well understood or explained. Managers, as well as their supervisors, peers, and subordinates, when providing their assessments may focus on a subset of CVF criteria that reflects the particular values that—from their vantage point—they believe are most relevant to successful managerial work.

There will be incongruence in 360-degree feedback ratings to the extent that individuals view managerial effectiveness differently and rely upon different frames of reference based on their roles. For example, the lower right CVF quadrant pertaining to goal achievement (identified herein as “improving results”) has long been recognized as the performance domain of priority for mid-level managers; in contrast, the upper right quadrant that maps the unique value placed on open systems (identified herein as “enhancing agility”) can require behaviors largely reserved for senior managers and, therefore, seem less relevant to employees at lower levels of an organizational hierarchy (Eden and Rohrbaugh 1990). Depending on the nature of the rater–ratee relationship, the degree of incongruence between rating sources can be expected to vary across the CVF quadrants, an issue that largely has been overlooked in previous studies. Following this reasoning and the preliminary evidence offered in recent studies (Hooijberg and Choi 2000; Pulkos et al. 1997; Salam et al. 1997), a third hypothesis was tested in the present study:

H3: The degree of incongruence between self–other and other–other ratings will vary in different performance domains.

All three hypotheses indicated above are based on prior research findings concerning incongruity in 360-degree feedback from studies of mostly private sector managers who varied considerably in their levels of professional performance (see, for example, Atwater et al. 1998; Johnson and Ferstl 1999; Van Velsor et al. 1993). The opportunity to investigate 360-degree feedback incorporated in a leadership training program for middle-level public sector managers allowed for a test of these hypotheses in a rather unique governmental context with participants specially selected for their demonstrated leadership potential. The current study anticipated that important and potentially informative rating incongruities (both other–other and self–other) would be found even for this select group of highly performing managers. Further, this study was proposed to demonstrate that explanations for such differences in assessments across performance domains would be attributable to the frames of reference distinctive of the evaluators’ roles, even under conditions in which these ratings would be expected to be somewhat uniformly high. This provided a particularly challenging and critical circumstance for the hypothesis tests that had been planned.

## RESEARCH METHOD

### Organizational Context

The data for this study were collected from a large local government organization (LGO)<sup>1</sup> located in the southwestern region of the United States. At the time of the study, the LGO had 37 departments and approximately 90,000 employees working in various areas of government including public protection, human services, recreation, and cultural services. A five-member board of supervisors served as both its executive and legislative body. The board was responsible for establishing and



enacting ordinances and administrative policies. A chief executive officer had the lead role in managing the LGO's financial stability and overseeing effective implementation of the board's decisions.

In the late 1990s, the LGO was facing a management crisis because of an aging workforce; approximately 30 percent of its managers were over the age of 55 and eligible for retirement in no more than a few years. In addition, due to budget constraints, many of its lower- and middle-level managers did not have access to leadership training. Consequently, they were not fully prepared to take over the responsibilities of their predecessors. With a relatively high turnover of managers, the LGO realized that it was no longer possible to meet leadership needs without a comprehensive educational program. Therefore, it made a new and substantial commitment to human resource development by investing a considerable amount of its resources to succession planning and management training.

The LGO worked very closely with a research team from a regional public university toward the establishment of a training academy that could offer a series of management development courses. The partnership between the LGO and the training academy was intended to improve managerial leadership capabilities in many ways, including the use of performance feedback; 360-degree feedback was one of several forms of performance feedback incorporated in the academy's programs. The primary goal of this initiative was to institute the idea that "feedback is a way of managerial life" to improve managers' performance. Another key goal was to address the resultant problems of its existing performance management system. As in other public organizations, the LGO was using a top-down approach to assess employees' performance. Many of its managers were dissatisfied with this system because they did not receive adequate performance feedback on a timely basis.

In particular, the LGO introduced 360-degree feedback in one of the training programs that targeted middle-level managers with demonstrated leadership potential who were eligible to apply for a position of division chief. To take part in this training program, trainees were required to be nominated by their supervisors. All of the trainees also were required to take a prerequisite test that assessed their preparedness for the training program. Each department determined the order in which trainees attended the program based on their job performance, the number of training slots allocated, the location of the training sessions, and the needs of the department. The training program consisted of 100 hours of classroom instruction over a 13-week period in various off-site training facilities. As part of the training, 360-degree feedback surveys were distributed to provide participants with additional information regarding their leadership potential. At the end of the training program, trainees received 360-degree feedback in the form of individualized reports. They also received one-on-one coaching from the university research team to help them in understanding and making further use of the 360-degree feedback.

### **Survey Design and Data Collection Process**

The research team at the university had extensive experience in consulting with a variety of public and private organizations on 360-degree feedback initiatives. They

used insights gained from those experiences in designing a survey and customizing it according to the LGO's unique organizational environment. They carefully reviewed LGO's mission, vision, and goals, and used the advice of its leaders to select items for the survey. Prior to administering the survey, a pilot initiative was undertaken to ensure both that the survey procedures would be effective and that the survey items would be relevant and clear. A group of trainees and their supervisors, subordinates, peers, and customers completed the pilot surveys. Later, in focus group discussion sessions, they provided detailed feedback about their experiences with the survey procedures and the usefulness of 360-degree feedback data to the research team. Based on the results of the pilot study, survey items were further evaluated and revised for the main initiative.

To collect data, trainees were asked to contact their supervisors, peers, subordinates, and customers and request them to provide a candid assessment of the trainee's performance according to several key managerial behavior and task dimensions. Each trainee also was asked to complete a self-evaluation of personal performance. A total of five cohorts of the leadership trainees were involved in this study. A total of 110 trainees, plus their supervisors, one or more peers, one or more subordinates, and in some cases their customers were asked to return completed surveys by mail. All of the 110 trainees completed self-assessments and 68 (62%) received feedback from their supervisors as well as one or more peers and subordinates; 42 (38%) trainees did not receive feedback from at least one of the three source of ratings (i.e., either from their supervisors or peers or subordinates) and were not included in this study. Of those 68 trainees who received feedback from all rating sources, 52% were male and 48% female.<sup>2</sup> As only a few trainees received customer feedback, those data are not included in the present study.

All respondents were informed about the intent of the performance evaluation process and how the feedback data would be used constructively in the training program. From the pilot to actual implementation to *post hoc* interviews, trainees were familiarized with the developmental objectives of the survey by the research team and instructors. Their peers, supervisors, subordinates, and customers also received similar communication from both the training academy and the LGO about the importance of the survey, as well as its intended use. To ensure confidentiality, all respondents were told that no one other than the trainees would receive a copy of their assessments. Additionally, peers, subordinates, and customers of each trainee were notified that their ratings would not be reported to the trainee unless two or more responses were received from the same rating source. Supervisors, however, did not receive such instruction because each trainee had only one supervisor. With frequent interactions among the research team, training academy, and the LGO, program instructors and LGO leaders effectively served as champions for the 360-degree feedback initiative. All these procedures were followed to enhance rater motivation and reduce rating biases (Bracken et al. 2001).

The final version of the survey was printed on four pages. The cover page specified the purpose of the survey, provided an assurance of confidentiality, asked respondents to identify their relationship with the trainee (supervisor, customer, direct report, or peer), and reaffirmed that responses would be aggregated for feedback

and leadership development. The second page explained the scales used for measuring the trainee's performance, so that each evaluator would be familiar with both the nature of overall scale and each rating. It then highlighted the LGO statement of its organizational goals of service excellence, fiscal integrity, organizational effectiveness, and workforce excellence. The next two pages presented 48 items designed to measure each rater's satisfaction with the way the trainee being rated undertook certain managerial tasks or exhibited certain behaviors. Each of the 48 questions was phrased initially as: "How satisfied are you with the way this individual...?" Each question was measured using a five-point scale from 1 = dissatisfied to 5 = highly satisfied. In addition, the survey provided space for comments and encouraged respondents to use additional sheets if needed. The two questions eliciting written feedback specifically asked: (1) what the individual being rated did most effectively and (2) what constructive suggestions could be provided.

Of the 48 items contained in the survey instrument, 12 were identified as being especially central to the quadrants of the CVF. Thus, the other 36 items were not included in this study. Only one survey from a peer and one survey from subordinate were randomly selected for each trainee.<sup>3</sup> To replace missing values in the data, the following rule was applied. If there was any alternative peer or subordinate evaluation with complete data to substitute for the one with missing data, then these data of the alternative peer or subordinate evaluation were used. If no substitute existed, then the missing item code was replaced with the arithmetic mean of the remaining items of the relevant scale. After replacing the missing values, the final data set consisted of a total of 272 surveys (68 from each rating source).

## RESULTS

Table 1 presents arithmetic means, standard deviations, and bivariate correlation coefficients of the 12 survey items used in this study. Trainees received fairly high performance ratings ( $M = 4.41$ , on a five-point scale) from all of the rating sources. The items for which trainees received the highest and lowest performance ratings were "*listens to ideas and suggestions of those doing work*" ( $M = 4.54$ ) and "*consistently researches new technologies to determine their applicability to service delivery*" ( $M = 4.19$ ), respectively. The standard deviations of all the items were relatively low on average ( $SD = 0.77$ ). The items with the highest and lowest variability in ratings were "*consistently researches new technologies to determine their applicability to service delivery*" ( $SD = .90$ ) and "*insures unit goals are clearly aligned with organization and department's goals*" ( $SD = .67$ ), respectively. All bivariate correlation coefficients were statistically significant ( $p < .05$ ). Further, correlations between items within the CVF quadrants were consistently higher than correlations with items in other quadrants.

Exploratory factor analysis was performed to organize the 12 survey items into the four dimensions of theoretical interest. This process involved identifying survey items that match the CVF framework and then examining their patterns with principal component analysis and varimax rotation. To check for consistency of the resultant model, factor analysis results of the combined data (272 surveys) were

**TABLE 1**  
Means, Standard Deviations, and Correlation Coefficients

<i>Behaviors</i>	<i>Mean</i>	<i>Std. Dev.</i>	<i>Supporting People</i>	<i>Enhancing Agility</i>	<i>Providing Stability</i>	<i>Improving Results</i>
Deals effectively with divergent points of view?	4.30	0.79				
Supports people in using their own initiative and creativity?	4.52	0.72	0.64			
Listens to the ideas and suggestions of those doing the work?	4.54	0.70	0.63 0.63			
Stays current with important technical changes in his or her field?	4.36	0.81	0.40 0.43 0.37			
Consistently researches new technologies to determine their applicability to service delivery?	4.19	0.90	0.40 0.42 0.39	0.79		
Participates in training and development programs to strengthen his or her technology skills?	4.45	0.79	0.41 0.40 0.40	0.78 0.74		
Insures unit goals are clearly aligned with organization and department's goals?	4.46	0.67	0.50 0.52 0.47	0.49 0.51 0.52		
Knows and uses all relevant organization's regulations and procedures to achieve unit goals?	4.45	0.71	0.40 0.37 0.40	0.45 0.48 0.42	0.60	
Effectively uses project management techniques to complete goals within established timelines and budgets?	4.25	0.89	0.52 0.50 0.49	0.50 0.50 0.49	0.58 0.57	
Is willing to change existing procedures and practices to improve results?	4.52	0.74	0.45 0.47 0.47	0.48 0.42 0.43	0.50 0.52 0.45	
Acknowledges results achieved as opposed to just following procedures?	4.42	0.70	0.54 0.51 0.47	0.50 0.50 0.49	0.59 0.45 0.53	0.70
Looks for new ways to improve service to customers?	4.45	0.76	0.46 0.56 0.50	0.51 0.49 0.49	0.46 0.46 0.47	0.73 0.73

compared with the factor analysis results of each separate source of ratings (68 surveys per source). A model with four factors that matched the CVF emerged clearly from this work. These four factors were: (1) supporting people for the human relations quadrant, (2) enhancing agility for the open systems quadrant, (3) providing stability for the internal processes quadrant, and (4) improving results for the rational goals quadrant of the CVF. These four factors together accounted for nearly 80 percent of the total variance, and each of them displayed the appropriately high factor loadings for its three designated items.

Table 2 presents results of the principal component analysis. Reliabilities (as shown with Cronbach's  $\alpha$ ) for all of the four factors were high: .84 for supporting people, .91 for enhancing agility, .80 for providing stability, and .88 for improving results. The factor analysis results were highly consistent both in the combined data and in each separate source of ratings. However, factor loadings from the combined data exhibited a somewhat simpler structure than factor loadings from separate sources of ratings. The consistency of high factor loadings (on average, .79) for all 12 items suggested that factor scores would be the most useful measure to conduct further analysis of rating incongruence. A set of four orthogonal factor scores<sup>4</sup> was generated from the combined data using the standard regression method to examine pattern of rating incongruence across rater groups; a check of the bivariate distributions confirmed that these factor scores were uncorrelated.<sup>5</sup>

Several methods are available to examine rating incongruence; difference score and correlational approaches seem the most intuitive ones. However, neither of these is an appropriate method to examine rating incongruence (Mersman and Donaldson 2000). Prior studies that have used a difference score approach experienced problems in correctly specifying the functional form of the relationships between self–other ratings and performance outcomes (Atwater et al. 1998; Edwards 1993a; 1993b; 1994; Johnson and Ferstl 1999; Mersman and Donaldson 2000). The correlational approach also is problematic because of the instability of correlation coefficients (Mersman and Donaldson 2000). Consistent with the approach adopted by Church (1997), this study relied on one-way repeated measures analysis of variance (ANOVA) with each of the four sets of factor scores as the dependent measure and rating source as the independent measure to examine the patterns of rating incongruence. The four analyses of variance were independent of each other because the factor scores were uncorrelated.

Table 3 presents a summary of the ANOVA results (by organizational roles). Mean performance ratings differed significantly ( $p < .05$ ) in three of the four CVF domains—enhancing agility ( $F = 15.67$ ), providing stability ( $F = 13.46$ ), improving results ( $F = 9.28$ )—by source of ratings. H1 proposed that there would be incongruence between self–other ratings. ANOVA results revealed that in two of the four CVF domains—enhancing agility and providing stability—trainee ratings differed significantly ( $p < .05$ ) from all other sources of ratings. In the CVF domain of improving results, trainee ratings differed significantly ( $p < .05$ ) from supervisor ratings. There were no significant differences in the CVF domain of supporting people. Thus, anticipated differences were found in three of the four CVF domains. These results provided partial empirical support for H1. In addition, the ANOVA results

**TABLE 2**  
Factor Pattern Matrices

Behaviors	Supporting People			Enhancing Agility			Providing Stability			Improving Results		
	All	Trainee		All	Trainee		All	Trainee		All	Trainee	
		Subordinate	Supervisor Peer		Subordinate	Supervisor Peer		Subordinate	Supervisor Peer		Subordinate	Supervisor Peer
Deals effectively with divergent points of view?	0.79	0.31 0.87	0.72 0.80	0.17	0.04 0.17	0.16 0.18	0.20	0.32 0.16	0.18 0.10	0.24	0.74 0.21	0.23 0.17
Supports people in using their own initiative and creativity?	0.78	0.65 0.82	0.29 0.85	0.21	0.15 0.23	0.16 0.17	0.29	0.25 0.18	0.23 0.12	0.15	0.44 0.27	0.72 0.16
Listens to the ideas and suggestions of those doing the work?	0.79	0.84 0.70	0.83 0.90	0.16	0.27 0.18	-0.05 0.12	0.23	0.24 0.23	0.08 0.07	0.21	0.13 0.52	0.11 0.07
Stays current with important technical changes in his or her field?	0.18	0.13 0.20	0.01 0.16	0.86	0.87 0.83	0.87 0.88	0.25	0.15 0.19	0.12 0.22	0.20	0.23 0.37	0.19 0.16
Consistently researches new technologies to determine their applicability to service delivery?	0.20	0.21 0.31	-0.06 0.27	0.85	0.82 0.81	0.88 0.85	0.21	0.29 0.27	0.15 0.07	0.20	0.19 -0.05	0.13 0.20
Participates in training and development programs to strengthen his or her technology skills?	0.19	0.30 0.11	0.22 0.13	0.83	0.71 0.84	0.86 0.90	0.19	0.36 0.18	-0.09 0.20	0.27	0.24 0.21	0.16 0.13
Insures unit goals are clearly aligned with organization and department goals?	0.33	0.15 0.24	0.47 0.46	0.29	0.21 0.26	0.04 0.23	0.68	0.68 0.84	0.32 0.64	0.25	0.36 0.01	0.52 0.29

Knows and uses all relevant agency regulations and procedures to achieve unit goals?	0.11	0.25	0.13	0.21	0.17	0.12	0.83	0.72	0.86	0.28	0.25	0.31
		-0.05	0.00		0.30	0.34		0.71	0.89		0.45	0.23
Effectively uses project management techniques to complete goals within established timelines and budgets?	0.40	0.17	0.52	0.30	0.30	0.30	0.65	0.79	0.26	0.17	0.08	0.40
		0.40	0.48	0.20	0.18	0.60	0.56	0.29	0.15			
Is willing to change existing procedures and practices to improve results?	0.21	0.23	-0.15	0.18	0.37	0.05	0.29	0.10	0.38	0.82	0.77	0.82
		0.27	0.05		0.13	0.16		0.56	0.53		0.72	0.85
Acknowledges results achieved as opposed to just following procedures?	0.31	0.04	0.41	0.26	0.44	0.13	0.28	0.36	-0.05	0.73	0.71	0.77
		0.48	0.26		0.24	0.07		0.17	0.16		0.64	0.85
Looks for new ways to improve service to customers?	0.31	0.48	0.19	0.28	0.45	0.18	0.14	0.17	-0.14	0.81	0.50	0.84
		0.36	0.14		0.18	0.25		0.14	0.09		0.85	0.88
Cronbach's $\alpha$	0.84			0.91			0.80			0.88		

**TABLE 3**  
One-Way Repeated Measures ANOVA Results (by Organizational Roles)

<i>Factors</i>	<i>Trainee</i>	<i>Supervisor</i>	<i>Peer</i>	<i>Subordinate</i>	<i>F Ratio</i>
Supporting people	-0.12	0.07	0.21	0.01	1.66
Enhancing agility	-0.49 <sup>a</sup>	0.15 <sup>b</sup>	0.05 <sup>b</sup>	0.24 <sup>b</sup>	15.67*
Providing stability	-0.58 <sup>a</sup>	0.02 <sup>b</sup>	0.26 <sup>b</sup>	0.28 <sup>b</sup>	13.46*
Improving results	-0.31 <sup>a</sup>	0.44 <sup>b</sup>	-0.16 <sup>a</sup>	0.00 <sup>a</sup>	9.28*

Note: Different superscripts indicate significant rating difference at  $p < .05$ . For example, supervisor ratings (b) of trainees' performance are different from all other (a) sources of ratings in the CVF domain of improving results.

\* $p < .05$ .

in Table 3 showed that trainees underrated their performance in two of the four CVF domains—enhancing agility and providing stability. In these two domains, trainees provided significantly lower ratings than their supervisors, peers, and subordinates and the mean factor scores were negative (enhancing agility = -.49, providing stability = -.58).<sup>6</sup>

H2 proposed that the degree of incongruence between self–other ratings would be higher than the incongruence between other–other ratings. As shown in Table 3, the results indicated no significant difference between supervisor–peer, supervisor–subordinate, and peer–subordinate ratings in three of the four CVF domains—supporting people, providing stability, and enhancing agility. In the CVF domain of improving results, supervisors did provide significantly ( $p < .05$ ) higher ratings than peers and subordinates. Thus, only two out of 12 possible differences were identified. These results provided clear empirical support for H2.

H3 proposed that the degree incongruence between self–other and other–other ratings would vary in different performance domains. To examine variation within each separate source of ratings, four one-way repeated measures ANOVA (by factors) were performed. The results as shown in Table 4 indicated that trainee

**TABLE 4**  
One-Way Repeated Measures ANOVA Results (by Factors)

<i>Factors</i>	<i>Trainee</i>	<i>Supervisor</i>	<i>Peer</i>	<i>Subordinate</i>
Supporting people	-0.12 <sup>b</sup>	0.07 <sup>a</sup>	0.21	0.01
Enhancing agility	-0.49 <sup>ab</sup>	0.15 <sup>ab</sup>	0.05	0.24
Providing stability	-0.58 <sup>a</sup>	0.02 <sup>a</sup>	0.26	0.28
Improving results	-0.31 <sup>ab</sup>	0.44 <sup>b</sup>	-0.16	0.00
F ratio	3.00*	3.49*	2.73	1.67

Note: Different superscripts indicate significant rating difference at  $p < .05$ . For example, supervisor ratings on trainee's performance in improving results (b) are different than their ratings in supporting people (a) and providing stability (a).

\* $p < .05$ .



and supervisor ratings differed significantly ( $p < .05$ ) across performance domains. *Post hoc* tests revealed that trainee ratings in the CVF domain of providing stability were significantly ( $p < .05$ ) lower than their self-ratings in the CVF domain of supporting people (providing stability–supporting people = .46). *Post hoc* tests also revealed that supervisor ratings of trainees in the CVF domain of improving results were significantly ( $p < .05$ ) higher from their ratings in both the CVF domains of providing stability and supporting people (improving results–providing stability = .46 and improving results–supporting people = .42).

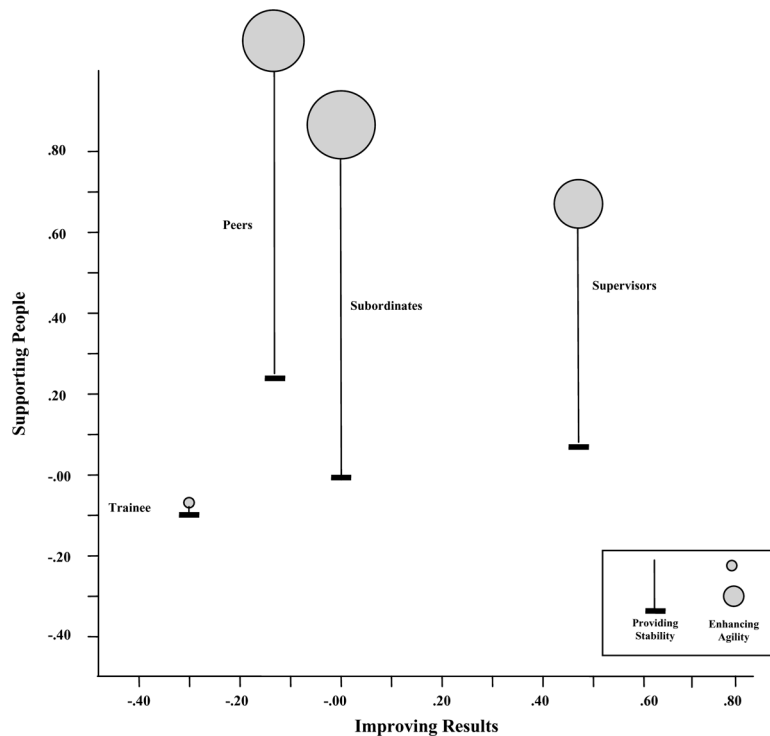
Figure 2 provides a full accounting of the extent of incongruence between trainee–other and other–other ratings across the four performance domains. As summarized in Table 3, the incongruence between trainee–other ratings was high both in the CVF domains of providing stability and enhancing agility due to low trainee self-ratings. The significant ( $p < .05$ ) mean rating differences in providing stability were the following: trainee–supervisor =  $-.61$ , trainee–peer =  $-.84$  and trainee–subordinate =  $-.86$ . The significant ( $p < .05$ ) mean rating differences in enhancing agility were the following: trainee–supervisor =  $-.65$ , trainee–peer =  $-.54$ , trainee–subordinate =  $-.73$ . The incongruence between trainee–other ratings and the incongruence between other–other ratings was high in the CVF domain of improving results due to the high ratings given to the trainees by their supervisors (see Table 4). The significant ( $p < .05$ ) mean rating differences in improving results were the following: trainee–supervisor =  $-.75$ , supervisor–peer =  $.60$ , and supervisor–subordinate =  $.45$ . Finally, the degree of incongruence between trainee–other and between other–other ratings was low in the CVF domain of supporting people. These findings provided strong empirical support for H3.

Figure 3 provides a graphic illustration of the degree of rating incongruence in the CVF domains as a four-dimensional space. The relative group positions of trainee self-ratings, peer ratings, subordinate ratings, and supervisor ratings are depicted as “balloons” spaced apart according to differences in their mean ratings for *supporting people* and mean ratings for *improving results* on the vertical and horizontal axes, respectively. The height of each “balloon” from its reference point indicates the

Trainee and Other Ratings		Other-Other Ratings	
Supporting People	Enhancing Agility	Supporting People	Enhancing Agility
Trainee-Supervisor = $-.19$ Trainee-Peer = $-.33$ Trainee-Subordinate = $-.13$	Trainee-Supervisor = $-.65^*$ Trainee-Peer = $-.54^*$ Trainee-Subordinate = $-.73^*$	Supervisor-Peer = $-.14$ Supervisor-Subordinate = $-.06$ Peer-Subordinate = $.20$	Supervisor-Peer = $.11$ Supervisor-Subordinate = $-.09$ Peer-Subordinate = $-.19$
Trainee-Supervisor = $-.61^*$ Trainee-Peer = $-.84^*$ Trainee-Subordinate = $-.86^*$	Trainee-Supervisor = $-.75^*$ Trainee-Peer = $-.15$ Trainee-Subordinate = $-.31$	Supervisor-Peer = $-.23$ Supervisor-Subordinate = $-.25$ Peer-Subordinate = $-.02$	Supervisor-Peer = $.60^*$ Supervisor-Subordinate = $.45^*$ Peer-Subordinate = $-.16$
Providing Stability	Improving Results	Providing Stability	Improving Results

\*Indicates significant mean differences at  $p < .05$  level

**Figure 2.** The Degree of Rating Incongruence Between Different Sources of Rating Across Performance Domains.



**Figure 3.** A Graphic Illustration of the Degree of Rating Incongruence in Four Performance Domains.

approximate magnitude of their mean ratings for *providing stability*; the size of each “balloon” indicates the approximate magnitude of their mean ratings for *enhancing agility*. Figure 3 illustrates clearly how supervisor ratings are set apart from the others primarily on the horizontal axis—improving results—and how trainee ratings are set apart from the others primarily in their small, short “balloons”—providing stability and enhancing agility. Peers and subordinates can be seen to occupy nearly the same location in this four-dimensional space.

## DISCUSSION AND CONCLUSIONS

This study examined the patterns of incongruence in 360-degree feedback ratings based on the Competing Values Framework (CVF) for organizational effectiveness. Statistical results provided supporting evidence for this conceptual model and, more importantly, for the three research hypotheses concerning rating incongruence in 360-degree feedback. Even within the select group of mid-level government managers who had been identified as prospective leaders, the ANOVA results (by organizational roles) of the present study were consistent with previous findings (Conway

and Huffcutt 1997; Furnham and Stringfield 1994; Harris and Schaubroeck 1988; Atwater and Yammarino 1992; Wohlers, Hall, and London 1993) that managers may assess their own performance differently than their peers, subordinates, and supervisors. Peer and subordinate ratings of managers' performance were found to be more closely related to each other than to the self-ratings of the managers.

The present study demonstrated that the degree of rating incongruence also can vary across different performance domains. For example, managers' self-ratings and their supervisors' ratings were found to converge or diverge depending on the CVF quadrant. This finding may encourage additional investigation into better approaches for designing and interpreting multi-level assessments. Furthermore, this finding indicates the presence of multiple sources of incongruence in 360-degree feedback ratings. Raters may differ in at least several ways of relevance to this issue: (a) in their perspectives on the components of managerial effectiveness (b) in the relative emphases that they place on multiple managerial roles (c) in the opportunities they have available to observe specific attributes of the managers whom they are asked to rate and (d) in the severity or leniency of the ratings that they produce. These explanations are consistent with Borman's arguments (1997) that raters may experience different managerial behaviors, employ different weight systems, and use different interpretive schemes in their evaluations.

That some raters may emphasize a particular subset of managerial roles more than others in 360-degree feedback ratings was evident in the supervisors' ratings of trainees. The ANOVA results (by organization roles) in Table 3 showed considerable rating incongruence in the CVF domain of improving results because ratings from supervisors were significantly higher than trainees' self, peer, and subordinate ratings. Further, the ANOVA results (by factors) in Table 4 showed that supervisor ratings of trainees in the CVF domain of improving results were significantly higher than their ratings both of providing stability and of supporting people. The significantly higher supervisor ratings in improving results may be attributed to the trainees' selection process in the LGO's management development program. The 68 trainees who participated in this study were a select group of high-performing, middle-level managers who had been nominated by their supervisors to participate in the management development program. Clearly, these trainees were very highly regarded by their supervisors for their success in improving results (the rational goals quadrant of the CVF) and, in all likelihood, were selected primarily for this reason.

The results of this study indicated that rating incongruence in the CVF domains of providing stability and enhancing agility was high because trainees provided significantly lower self-ratings than their peer, supervisor, and subordinate ratings. There are several possible explanations for why trainees decreased their self-ratings in these two performance domains. One possibility is that these high-performing trainees underestimated their performance because they were not fully aware of their ability compared to others (Burson et al. 2006; Kruger and Dunning 1999; 2002). Another explanation is that trainees modestly underestimated aspects of their performance to avoid an appearance of self-aggrandizement. This could result from uncertainty about the nature of the scrutiny that would be given to 360-degree feedback in the training program. Although efforts were taken in the data collection process to

reduce such concerns, trainees had little prior experience with 360-degree feedback evaluation methods, possibly inducing cautiousness in the self-rating process.

Another, perhaps more plausible, explanation consistent with CVF theory is that trainees underrated their performance because they believed they were somewhat organizationally constrained by their current position in the hierarchy—due to agency rules and procedures, as well as the allocation of resources—in providing stability and enhancing agility. In short, the trainees may be understandably more critical about themselves with respect to those CVF quadrants in which they are least likely to bear the burden of responsibility. In any case, previous studies have found that it is low-performing—not high-performing—managers who tend to produce inflated self-ratings due to a variety of cognitive factors (Atwater and Yammarino 1992; Furnham and Stringfield 1994). This source of incongruity primarily manifests when 360-degree feedback is used for performance appraisal rather than in the present context of management development.

Atwater and others (1998) have reported that the effectiveness of managers was highest when both managers' self and other ratings were all fairly high, and when managers' self-ratings were lower than the ratings of others. As indicated by the ratings of their supervisors, peers, and subordinates, selected trainees in this study already had been judged to be highly effective, even though they underrated their own performance in two of the four CVF domains—enhancing agility and providing stability. Some researchers have suggested that underraters tend to be effective and successful managers, as they have a tendency to overestimate their weaknesses and underestimate their strengths (Ashford 1989; Atwater et al. 1998; Atwater, Roush, and Fischthal 1995; Atwater and Yammarino 1997; Yammarino and Atwater 1997). Underraters also are thought to be more open to negative feedback than others (Ashford 1989). The tendency to overestimate their weaknesses perhaps motivates underraters to work hard and, in turn, may result in greater success at tasks they undertake (Atwater et al. 1998). Mersman and Donaldson (2000), however, have suggested that, because underraters may exhibit a lack of self confidence, they appear to have difficulty in making important decisions regarding their future career goals and objectives.

Whereas trainees' self-ratings diverged from the ratings of others in the CVF domains of providing stability and enhancing agility, they were convergent with the ratings of others in the CVF domains of supporting people and improving results.<sup>7</sup> The ANOVA results (by organizational roles) in Table 3 indicated that rating incongruence in the CVF domain of supporting people was low because trainees' self-ratings were close to their supervisor, peer, and subordinate ratings. The survey items that loaded highly on the factor of supporting people assessed the trainees' performance in important interpersonal behaviors that are essential to building strong work relationships. To be highly effective, the trainees' performance required a great degree of shared experience. Thus, convergence of trainees' self-ratings with their peer, supervisor, and subordinate ratings in supporting people might be expected. The ANOVA results (by organizational roles) in Table 3 also indicated that incongruence between trainees' self and their peer ratings and incongruence between trainees' self and their subordinate ratings were both low; differences were not

statistically significant. The survey items in the CVF domain of improving results mainly captured the trainees' performance in task-related behaviors. Since task-related behaviors are observed readily and often quantified and linked to unit outcomes, compared to behaviors in other performance domains (Bales 1950), the convergence of trainees' self-ratings with their peer and subordinate ratings in improving results would seem to be a valid finding of this study, as well.

Consistent with the research of Atwater et al. (1998) and Johnson and Ferstl (1999), the results of this study indicated that it is important to focus on both the magnitude and the direction of managers' self and other ratings to interpret the patterns of incongruence in 360-degree feedback ratings. The variation in the degree of incongruence among different rating sources and across different performance domains found in this study points to the need for adopting an appropriate statistical method for examining rating incongruence. Correlations between raters are a poor measure of rating incongruence, for example, because these coefficients are insensitive to even large differences between two measures' central tendencies and/or dispersions. Had this study used either a difference score or correlational approach to examine rating incongruence, key observations concerning the degree and direction of rating incongruence would not have been so evident. The innovative use of a repeated measures ANOVA design as illustrated in this study is arguably more appropriate for the type of hypothesis tests that can resolve major theoretical issues concerning rating incongruence in 360-degree feedback.

One of the most attractive features of 360-degree feedback is that it gives an initial impression of objectivity and fairness in the performance evaluation process (Toegel and Conger 2003). Therefore, one may see a clear advantage in using 360-degree feedback either as a management development or as a performance appraisal tool. However, as mentioned above, extending the use of 360-degree feedback for potentially conflicting organizational objectives creates the risk of serious negative consequences for managers. If an agency were to use 360-degree feedback for making decisions on managers' compensation, promotion, and termination, it typically would prefer to minimize rating incongruence and seek a consistent measure of managers' overall performance (Toegel and Conger 2003) for increased validity. However, the primary purpose for the use of 360-degree feedback is human resource development, by providing managers with candid information about their performance from multiple perspectives, even if these perspectives are somewhat contradictory. As shown in the present study, the fact that different perspectives in rating sources and across different performance domains can lead to rating incongruence is not necessarily a problem but good evidence that organizational actors have different goal expectations. Thus, incongruence in 360-degree feedback ratings can reflect legitimate differences in the raters' perceptions in at least four distinct performance domains.

Both researchers and practitioners who are interested in the use of 360-degree feedback for management development in public organizations can draw several useful lessons from this study. First, it is important to consider all sources of input when providing managers with feedback and guidance. Even though this study did not find significant differences between peer and subordinate ratings of the trainees, this

does not necessarily mean that they are substitutable or unnecessary. In alternative organizational contexts, their assessments might be expected to diverge. Excessive reliance on any particular source of ratings—for example, only supervisors—would be much less useful than the potential variety of insights provided by multiple rating sources. Second, differences in the perspectives of raters should be made explicit when providing guidance to managers for their professional development. It is important that public sector managers be aware that those with whom they work may have different goal expectations and that they may provide their evaluations using different priorities for managerial effectiveness. Third, public sector managers should carefully consider the patterns of congruence and incongruence between their own ratings and the ratings of their supervisors, peers, and subordinates in different performance domains while interpreting feedback. Such insight will assist them in assessing how well they are meeting the needs of different stakeholders, as well as in taking necessary steps toward further improvement of their performance. Finally, for 360-degree feedback to be effective in initiating meaningful change in public organizations, it is important that public sector managers are provided with appropriate training about how to analyze, interpret, and use 360-degree feedback constructively. Further, public sector managers should devise a professional development plan that aligns their individual performance objectives and career expectations with organizational goals and succession planning.

One might argue that a limitation of the present research design that the trainees were a select group of high performing middle-level managers with demonstrated leadership potential. Clearly, the sizeable item means and small item standard deviations are indicative of a potential statistical problem of “restricted ranges,” that is, the possibility that expected relationships between variables across a heterogeneous population may be attenuated in a more homogeneous sample. Where correlational analyses were undertaken in the present study, however, anticipated connections between items were strong and significant, despite restricted ranges. However, one might conclude that the results reported here are conservative because they emerged even within a “restricted range” condition. Greater differentiation, when possible to achieve, in all likelihood would only more starkly underscore the major points of the present study. Furthermore, it is important to note that one should be cautious in generalizing the findings of this study to all managers or even all middle-level managers in public agencies, especially the expectation that self-assessments will be incongruent with other assessments due to the likelihood of managers’ underrating.

Another limitation of this study that should be noted was its design implication that the different values held by organizational stakeholders lead to contrasting views about managerial effectiveness, which, in turn, are reflected in the incongruent performance assessments of the trainees, supervisors, peers, and subordinates. Clearly, this is a highly functionalist approach. It assumes that individuals’ values are determined largely by their own organizational roles. Admittedly, values also can be influenced by a variety of personal and interpersonal variables, as well as by larger social, political, economic, and cultural conditions. Due to constraints placed upon the present study by limitations of available data, only the effect of differences in organizational roles on the patterns of rating incongruence could be

addressed. Hence, future research on 360-degree feedback should be encouraged to investigate the extent to which many other factors also can influence performance assessments.

This study also was unable to examine the relationship between rating incongruence, managerial performance, and organizational outcomes over time due to the lack of necessary data. A number of previous studies have investigated the performance implications of incongruence in 360-degree feedback ratings. Some of these (Atwater and Yammarino 1992; Bass and Yammarino 1991; Feinberg, Ostroff, and Burke 2005; Furnham and Stringfield 1994; Van Velsor et al. 1993), for example, have suggested that the performance of overraters tends to be lower than that of underraters. However, empirical evidence concerning the longitudinal impact of 360-degree feedback on the performance of underraters remains inconclusive (Bailey and Fletcher 2002; Smither et al. 1995) and requires further investigation. Thus, an important and interesting future research question is whether underraters change their behaviors and job performance after they receive feedback from their supervisors, peers, and subordinates.

360-degree feedback can play an important role in public managers' career learning and development. The increased popularity of all forms of multi-source feedback among public sector organizations, in fact, is a clear demonstration of the desire of public managers to improve their performance and increase organizational effectiveness. An obvious deficiency exists in the research literature, however, regarding the contingencies of multi-source feedback use in public agency settings. In comparison to private organizations, public organizations tend to favor higher degrees of formalization with clear division of labor, and well-defined departmentalization coupled with stable hierarchical arrangements (Boyne 2002; Bozeman, Reed, and Scott 1992; Kurland and Egan 1999; Perry and Rainey 1988; Rainey 1983; Rainey and Bozeman 2000; Rainey, Pandey, and Bozeman 1995). These general structural differences are the result of both an appropriate response to environmental demands (Child 1972; Thompson 1967), as well as adaptation to institutional norms (DiMaggio and Powell 1983). Nevertheless, they may have important implications in the use and consequences of 360-degree feedback for management development in public organizations. Over the past three decades, considerable pressure has been placed upon public managers to adopt techniques and practices that are popular for human resource management in private sector organizations (Boyne 2002; Golembiewski 1995). However, one must carefully consider possible differences in tasks, work context, and organizational structure to assess whether any management development technique, including 360-degree feedback, has a useful function in human resource management for a particular agency.

The most common approach to performance assessment in public organizations is the one in which only supervisors rate their subordinates' performance. The results of this study, however, have highlighted the value for including multiple perspectives in the performance assessment process in public organizations. For instance, supervisor ratings were significantly different from peer and subordinate ratings on trainees' performance in the CVF domain of improving results. Additionally, both the trainees themselves and their supervisors had varying perspectives on the

trainees' effectiveness across the four performance domains. Of course, these differences do not necessarily mean that either the trainees or their supervisors were incorrect in their assessments, but they do underscore that incongruence is expected and valuable in 360-degree ratings. The efficacy of 360-degree feedback will depend upon understanding such differences and interpreting them wisely for management development to occur in public organizations.

## ACKNOWLEDGEMENTS

The authors are especially grateful to Alan M. Glassman, Mingfang Li, Gerard Rossy, Kenneth Chapman, and Liz Barrett for providing support and guidance in carrying out this research study. The authors also appreciate the helpful comments of the two anonymous reviewers. An earlier version of this paper was presented at the eighth bi-annual Public Management Research Conference in Tucson, Arizona.

## NOTES

1. The LGO operated on an annual budget of more than 20 billion dollars.
2. The survey did not include any questions regarding the raters' demographic characteristics or other organizational information. Therefore, raters' demographic and organizational data cannot be summarized in this paper.
3. The intent of this approach was to adhere to a strictly individual-level analysis. This ameliorated concern that averaging the ratings of two or more peers (or subordinates) would introduce incommensurate "group-level" statistics as data. Additionally, although an average might reduce error variability in estimation, in this context it also could obscure important individual differences in perception that were the focus of this study.
4. Factor scores were computed in the following way. For each respondent, the measurement on each item (in standard form) was multiplied by its respective pattern weight. The sum of these weight-times-measurement products for all 12 items yielded a respondent's factor score. Resulting factor scores are presented in standard form with mean of 0.0 and standard deviation of 1.0. Approximately two-thirds of all factor scores can be found between +1.0 and -1.0.
5. Combined data were used in generating the factor scores to assure maximum comparability in scaling; otherwise, differences could be attributed, in part, to scaling (i.e., if factor scores had been generated uniquely for each of the four sources of ratings).
6. Parallel ANOVA results produced by relaxing the repeated measures assumptions and using an independent treatment model yielded the same pattern of statistically significant differences. A MANOVA approach was not indicated because the four dependent measures were derived from orthogonal factors.
7. The exception was between trainees' self and supervisor ratings where greater incongruence was due to high supervisor ratings.

## REFERENCES

- Ashford, S. 1989. "Self-Assessments in Organizations: A Literature Review and Integrated Model." *Research in Organizational Behavior* 11: 133-174.



- Ashford, S. and A. Tsui. 1991. "Self-Regulation for Managerial Effectiveness: The Role of Active Feedback Seeking." *Academy of Management Journal* 34: 251–280.
- Atkins, P. W. and R. E. Wood. 2002. "Self-Versus-Others' Ratings as Predictors of Assessment Center Ratings: Validation Evidence for 360-degree Feedback Programs." *Personnel Psychology* 55: 871–904.
- Atwater, L. E., C. Ostroff, F. J. Yammarino, and J. W. Fleenor. 1998. "Self-Other Agreement: Does It Really Matter?" *Personnel Psychology* 51: 577–598.
- Atwater, L. E., P. Roush, and A. Fischthal. 1995. "Does Self-Other Agreement on Leadership Perceptions Moderate the Validity of Leadership and Performance Predictions?" *Personnel Psychology* 45: 141–164.
- Atwater, L. E., D. A. Waldman, C. Robie, C. Ostroff, and J. Johnson. 2005. "Self-Other Agreement: Comparing Its Relationship with Performance in U.S. and Europe." *International Journal of Selection and Assessment* 13: 25–40.
- Atwater, L. E. and F. J. Yammarino. 1992. "Does Self-Other Agreement on Leadership Perceptions Moderate the Validity of Leadership and Performance Predictions?" *Personnel Psychology* 45: 141–164.
- Atwater, L. E. and F. J. Yammarino. 1997. "Self-Other Rating Agreement: A Review of the Model." *Research in Personnel and Human Resource Management* 15: 121–174.
- Bailey, C. and C. Fletcher. 2002. "The Impact of Multiple Source Feedback on Management Development: Findings from a Longitudinal Study." *Journal of Organizational Behavior* 23: 853–867.
- Bales, R. F. 1950. *Interaction Process Analysis: A Method for the Study of Small Groups*. Chicago: University of Chicago Press.
- Bass, B. and F. J. Yammarino. 1991. "Congruence of Self and Others' Leadership Ratings of Naval Officers for Understanding Successful Performance." *Applied Psychology: An International Review* 40: 437–454.
- Borman, W. C. 1997. "360 Degree Ratings: An Analysis of Assumptions and a Research Agenda for Evaluating their Validity." *Human Resource Management Review* 7: 299–315.
- Boyne, G. A. 2002. "Public and Private Management: What's the Difference?" *Journal of Management Studies* 39: 97–122.
- Bozeman, B., P. Reed, and P. Scott. 1992. "Red Tape and Task Delays in Public and Private Organizations." *Administration and Society* 24: 290–322.
- Bozeman, D. 1997. "Interrater Agreement in Multi-Source Performance Appraisal: A Commentary." *Journal of Organizational Behavior* 18: 213–316.
- Bracken, D., C. W. Timmreck, and A. H. Church. 2001. *The Handbook of Multisource Feedback: The Comprehensive Resource for Designing and Implementing MSF Processes*. San Francisco: Jossey-Bass.
- Burson, K. A., R. P. Larrick, and J. Klayman. 2006. "Skilled or Unskilled, but Still Unaware of It: How Perceptions of Difficulty Drive Miscalibration in Relative Comparisons." *Journal of Personal and Social Psychology* 90: 60–77.
- Child, J. 1972. "Organization Structure, Environment, and Performance: The Role of Strategic Choice." *Sociology* 6: 1–22.
- Church, A. H. 1997. "Do You See What I See? An Exploration of Congruence in Ratings from Multiple Perspectives." *Journal of Applied Social Psychology* 27: 883.
- Conway, J. M. and A. E. Huffcutt. 1997. "Psychometric Properties of Multi-Source Feedback Ratings: A Meta-Analysis of Subordinate, Supervisor, Peer and Self-Ratings." *Human Performance* 10: 331–360.
- Dalton, M. 1997. "When the Purpose of Using Multi-Rater Feedback Is Behavior Change." Pp. 1–6 in D. Bracken, M. Dalton, R. Jako, C. McCarthy, and V. Pollman, eds., *Should*

- 360-Feedback Be Used Only for Development Purposes?* Greensboro, NC: Center for Creative Leadership.
- deLeon, L. and A. J. Ewen. 1997. "Multi-Source Performance Appraisals." *Review of Public Personnel Administration* 17: 22–36.
- DeNisi, A. and K. Williams. 2000. "Feedback Effectiveness: Can 360-Degree Appraisals Be Improved?" *Academy of Management Executive* 14: 129–139.
- Denison, D. R., R. Hooijberg, and R. E. Quinn. 1995. "Paradox and Performance: Toward a Theory of Behavioral Complexity in Managerial Leadership." *Organization Science* 6: 520–540.
- DeVader, C. L., A. G. Bateson, and R. G. Lord. 1986. "Attribution Theory: A Meta-Analysis of Attributional Hypothesis." Pp. 63–81 in E. Locke, ed., *Generalizing from Laboratory to Field Settings*. Lexington, MA: Lexington Books.
- DiMaggio, P. J. and W. W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48: 147–160.
- Eden, C. F. and J. Rohrbaugh. 1990. "Using the Competing Values Approach to Explore 'Ways of Working.'" Pp. 40–47 in C. F. Eden and J. Radford, eds., *Tackling Strategic Problems: The Role of Group Decision Support*. Newbury Park, CA: Sage.
- Edwards, J. R. 1993a. "On the Use of Polynomial Regression Equations as an Alternative to Difference Scores in Organizational Research." *Academy of Management Journal* 36: 1577–1613.
- Edwards, J. R. 1993b. "Problems with the Use of Profile Similarity Indices in the Study of Congruence in Organizational Research." *Personnel Psychology* 46: 641–665.
- Edwards, J. R. 1994. "The Study of Congruence in Organizational Behavior Research: Critique and Proposed Alternative." *Organizational Behavior and Human Decision Processes* 58: 683–689.
- Fahr, J. and G. Dobbins. 1989. "Effects of Self-Esteem on Leniency Bias in Self-Reports of Performance: A Structural Equation Model Analysis." *Personnel Psychology* 42: 835–850.
- Fayol, H. 1916. *Administration Industrielle et Générale*. Paris: Dunod.
- Feinberg, B., C. Ostroff, and W. Burke. 2005. "The Role of Within-Group Agreement in Understanding Transformational Leadership." *Journal of Occupational and Organizational Psychology* 78: 471–488.
- Fox, S. and Y. Dinur. 1988. "Validity of Self-Assessment: A Field Evaluation." *Personnel Psychology* 41: 581–592.
- Furnham, A. and P. Stringfield. 1994. "Congruence in Self and Subordinate Ratings of Managerial Practices as a Correlate of Supervisor Evaluation." *Journal of Occupational and Organizational Psychology* 67: 57–67.
- Golembiewski, R.T. 1995. *Practical Public Management, Public Administration and Public Policy*. New York: M. Dekker.
- Gulick, L. 1937. "Notes on the Theory of Organizations. With Special Reference to Government." Pp. 1–46 in L. Gulick and L. Urwick, eds., *Papers on the Science of Administration*. New York: Institute of Public Administration, Columbia University.
- Harris, M. and J. Schaubroeck. 1988. "A Meta-Analysis of Self-Supervisor, Self-Peer, and Peer-Supervisor Ratings." *Personnel Psychology* 41: 43–61.
- Hooft, E. A., F. Henk, and M. R. Minne. 2006. "Construct Validity of Multi-Source Performance Ratings: An Examination of the Relationship of Self-Supervisor, and Peer-Ratings with Cognitive and Personality Measures." *International Journal of Selection and Assessment* 14: 67–81.

- Hooijberg, R. and J. Choi. 2000. "Which Leadership Roles Matter to Whom? An Examination of Rater Effects on Perceptions of Effectiveness." *Leadership Quarterly* 11: 341–364.
- Johnson, J. and K. Ferstl. 1999. "The Effects of Interrater and Self–Other Agreement on Performance Improvement Following Upward Feedback." *Personnel Psychology* 52: 271–303.
- Kellough, J. E. and S. C. Selden. 1997. "Pay-for-Performance Systems in State Government: Perception of State Agency Personnel Managers." *Review of Public Personnel Administration* 17: 5–21.
- Krugar, J. and D. Dunning. 1999. "Unskilled and Unaware of It: How Difficulties in Recognizing Ones' Own Incompetence Lead to Inflated Self-Assessments." *Journal of Personality and Social Psychology* 77: 1121–1134.
- Krugar, J. and D. Dunning. 2002. "Unskilled and Unaware—But Why? A Reply to Krueger and Mueller." *Journal of Personality and Social Psychology* 82: 189–192.
- Kurland, N. and T. Egan. 1999. "Public v. Private Perceptions of Formalization, Outcomes, and Justice." *Journal of Public Administration Research and Theory* 9: 437–458.
- Locke, E. A. 1968. "The Effects of Knowledge of Results, Feedback in Relation to Standards, and Goals on Reaction Time Performance." *American Journal of Psychology* 81: 566–574.
- Locke, E. A. and G. P. Latham. 1990. *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice Hall.
- McCartt, A. T. and J. Rohrbaugh. 1995. "Managerial Openness to Change and the Introduction of GDSS: Explaining Initial Success and Failure in Decision Conferencing." *Organization Science* 6: 569–584.
- Mersman, J. and S. Donaldson. 2000. "Factors Affecting the Convergence of Self–Peer Ratings in Contextual and Task Performance." *Human Performance* 13: 299–322.
- Miller, J. S. and R. L. Cardy. 2002. "Self-Monitoring and Performance Appraisal: Rating Outcomes in Project Teams." *Journal of Organizational Behavior* 21: 609–626.
- Mintzberg, H. 1973. *The Nature of Managerial Work*. New York: Harper & Row.
- Mount, M. K., T. A. Judge, S. E. Scullen, M. R. Sytsma, and S. R. Hezlett. 1998. "Trait, Rater and Level Effects in 360-Degree Performance Ratings." *Personnel Psychology* 51: 557–576.
- Murphy, K. R., and J. N. Cleveland. 1995. *Understanding Performance Appraisal: Social, Organizational, and Goal-Based Perspectives*. Thousand Oaks, CA: Sage.
- Pandey, S. K. and B. E. Wright. 2006. "Connecting the Dots in Public Management: Political Environment, Organizational Goal Ambiguity, and the Public Manager's Role Ambiguity." *Journal of Public Administration Research and Theory* 16: 511–532.
- Parson, T. 1959. "General Theory in Sociology." Pp. 3–38 in R. Merton, L. Broom, and L. S. Cottrell, Jr., eds., *Sociology Today: Problems and Prospects*. New York: Basic Books.
- Perry, J. L. and H. G. Rainey. 1988. "The Public–Private Distinction in Organization Theory: A Critique and Research Strategy." *Academy of Management Review* 13: 182–201.
- Podsakoff, P. M. and D. W. Organ. 1986. "Self-Reports in Organizational Research: Problems and Prospects." *Journal of Management* 12: 531–544.
- Pollack, D. and L. Pollack. 1996. "Using 360-Degree Feedback in Performance Appraisal." *Public Personnel Management*, 25: 507–528.
- Pollman, V. 1997. "Some Faulty Assumptions That Support Using Multi-Rater Feedback for Performance Appraisal." Pp. 7–9 in D. Bracken, M. Dalton, R. Jako, C. McCarthy, and V. Pollman, eds., *Should 360-Feedback Be Used Only for Development Purposes?* Greensboro, NC: Center for Creative Leadership.

- Pulakos, E. D., N. Schmitt, and D. Chan. 1996. "Models of Job Performance Ratings: An Examination of Ratee Race, Ratee Gender, and Rater Level Effects." *Human Performance* 9: 103–119.
- Pynes, J. E. 1997. *Human Resource Management in for Public and Nonprofit Organizations*. San Francisco: Jossey-Bass.
- Quinn, R. E. and J. Rohrbaugh. 1983. "A Spatial Model of Effectiveness Criteria: Towards a Competing Values Approach to Organizational Analysis." *Management Science* 29: 363–377.
- Rainey, H. G. 1983. "Public Agencies and Private Firms: Incentive Structures, Goals, and Individual Roles." *Administration and Society* 15: 65–105.
- Rainey, H. G. 1989. "Public Management: Recent Research on the Political Context and Managerial Roles, Structures, and Behaviors." *Journal of Management* 15: 229–50.
- Rainey, H. G. 1997. *Understanding and Managing Public Organizations*. San Francisco: Jossey-Bass.
- Rainey, H. G. and B. Bozeman. 2000. "Comparing Public and Private Organizations: Empirical Research and the Power of the A Priori." *Journal of Public Administration Research and Theory* 10: 447–469.
- Rainey, H. G., S. Pandey, and B. Bozeman. 1995. "Research Note: Public and Private Managers' Perceptions of Red Tape." *Public Administration Review* 55: 567–574.
- Rohrbaugh, J. 1983. "The Competing Values Approach: Innovation and Effectiveness in the Job Service." Pp. 165–280 in R. H. Hall and R. E. Quinn, eds., *Organization Theory and Public Policy*. Beverly Hills: Sage.
- Rohrbaugh, J. 2005. "Competing Values: Assessing the Effectiveness of Group Facilitation." Pp. 449–456 in S. Schuman, ed., *The IAF Group Facilitation Handbook*. San Francisco: Jossey-Bass.
- Salam, S., J. Cox, and H. Sims. 1997. "In the Eye of the Beholder: How Leadership Relates to 360-Degree Performance Ratings." *Group and Organization Management* 22: 185–209.
- Selden, S. C., P. I. Ingraham, and W. Jacobson. 2001. "Human Resource Practices in State Government: Findings from a National Survey." *Public Administration Review* 61: 598–607.
- Smither, J. W., M. London, N. L. Vasilopoulos, R. R. Reilly, R. E. Millsap, and N. Salvemini. 1995. "An Estimation of the Effects of Upward Feedback Program over Time." *Personnel Psychology* 48: 1–34.
- Thompson, J. D. 1967. *Organizations in Action*. New York: McGraw-Hill.
- Toegel, G. and J. A. Conger. 2003. "360-Degree Assessment: Time for Reinvention." *Academy of Management Learning and Education* 2: 297–311.
- Tornow, W. W. 1993. "Perceptions or Reality: Is Multi-perspective Measurement a Means or an End?" *Human Resource Management* 32: 221–229.
- Vandenberg, R. J., C. E. Lance, and S. C. Taylor. 1997. "A Latent Variable Approach to Rating Sources Equivalence: Who Should Provide Ratings on Organizational Citizenship Behavior Dimensions?" Academy of Management Annual Conference, Boston, MA, August 11–14.
- Van Velsor, E., S. Taylor, and J. Leslie. 1993. "An Examination of the Relationships Among Self-Perception Accuracy, Self-Awareness, Gender, and Leader Effectiveness." *Human Resource Management* 32: 249–264.
- Waldman, D. A., L. E. Atwater, and D. Antonioni. 1998. "Has 360-Feedback Gone Amok?" *Academy of Management Executive* 12: 86–94.
- Warr, P. and A. Bourne. 1999. "Factors Influencing Two Types of Congruence in Multi-Rater Judgments." *Human Performance* 12: 183–210.

- Wohlers, A. J., M. J. Hall, and M. London. 1993. "Subordinates Rating Managers: Organizational and Demographic Correlates of Self/Subordinate Agreement." *Journal of Occupational and Organizational Psychology* 66: 263–275.
- Wohlers, A. J. and M. London. 1989. "Ratings of Managerial Characteristics: Evaluation Difficulty, Co-Worker Agreement, and Self-Awareness." *Personnel Psychology* 42: 235–261.
- Yammarino, F. J. and L. E. Atwater. 1993. "Understanding Self-Perception Accuracy: Implications for Human Resource Management." *Human Resource Management* 32: 231–247.
- Yammarino, F. J. and L. E. Atwater. 1997. "Do Managers See Themselves as Others See Them? Implications of Self-Other Rating Agreement for Human Resources Management." *Organizational Dynamics* 25(4): 35–44.

### ABOUT THE AUTHORS

**Shahidul Hassan** (sh751124@albany.edu) is a Ph.D. student at the Rockefeller College of Public Affairs and Policy at the University at Albany (SUNY). His research interests include ethical issues, justice, and fairness in organization, managerial competence, organizational identification, and commitment.

**John Rohrbaugh** (jwr26@albany.edu) earned his Ph.D. in social psychology at the University of Colorado and currently serves as professor at the Rockefeller College of Public Affairs and Policy at the University at Albany (SUNY). His research has focused on the problem-solving processes of management groups, executive teams, and expert task forces to identify methods that would improve organizational decision making.

Copyright of *International Public Management Journal* is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.